IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ali Yafi

14-August-2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Due to High cost from SpaceX Experiment, SpaceY want to launch rocket with less cost by predicting first stage launch. Data Collected from SpaceX API and Wiki URL then wrangled and EDA to find hidden pattern of dataset. Interactive Map & predicting models were created to give better insight & to predict first stage of next launch.

SpaceY can reduce cost by predicting first stage of launching with accuracy 83.3% using Decision Tree Classifier with 4 Launch site used by SpaceX with the highest success rate is KSC LC-39A. Success Rate of Rocket Launch increasing year by year, so SpaceY can adapt concept from SpaceX.

Introduction

Context

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

Background

Due to High cost from SpaceX Experiment, SpaceY want to launch rocket with less cost.

Problem

How SpaceY reuse the first stage of rocket launch to reduce cost?

Section 1

Methodology

Methodology

Executive Summary

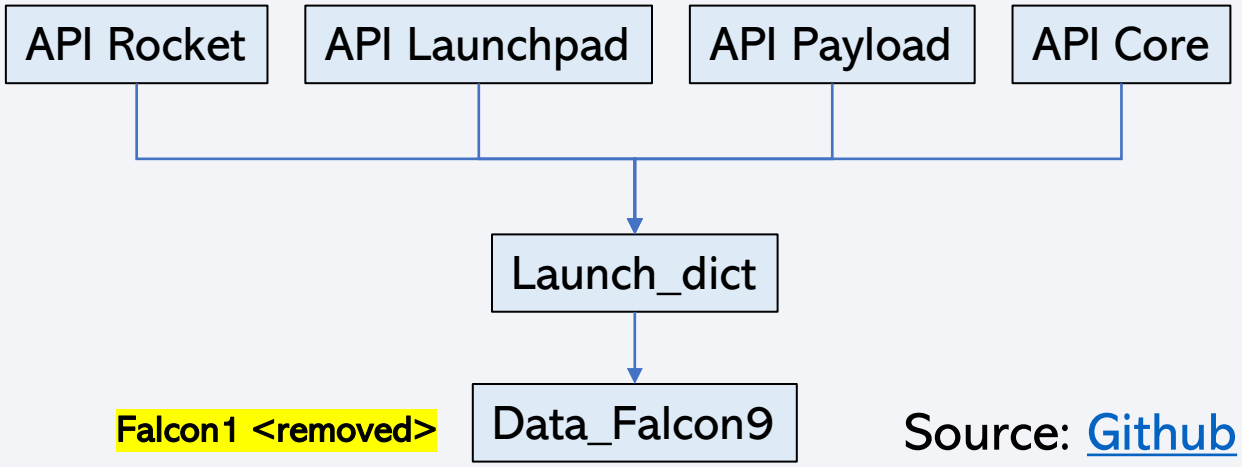
- Data collection methodology:
  - SpaceY Collect Data from Webscapping and SpaceX API
- Perform data wrangling
  - Handling Missing Value and Create Class each obeservations
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Split data into train and test, tune by GridSearchCV, Evaluate by Score

Data Collection

- Data was collected by Webscapping and REST API method
- Webscrapping link: https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- REST API:
  - https://api.spacexdata.com/v4/rockets/
  - https://api.spacexdata.com/v4/launchpads/
  - https://api.spacexdata.com/v4/payloads/
  - https://api.spacexdata.com/v4/cores/
  - https://api.spacexdata.com/v4/launches/past
  - Json: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json

Data Collection – SpaceX API

This Process to collect and make sure the data is in the correct format from an API



Falcon1 <removed>

Source: Github

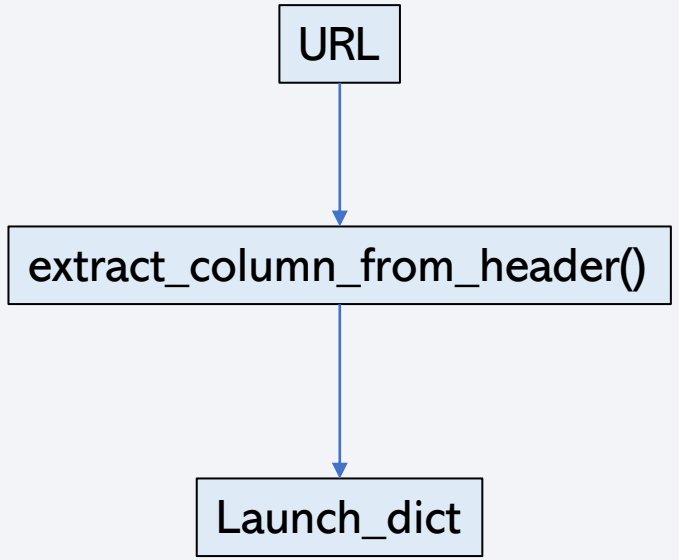
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	8	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	10	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	11	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	12	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
89	102	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1060	-80.603956	28.608058
90	103	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	13	B1058	-80.603956	28.608058
91	104	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1051	-80.603956	28.608058
92	105	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	12	B1060	-80.577366	28.561857
93	106	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	8	B1062	-80.577366	28.561857

90 rows × 17 columns

Data Collection - Scraping

This Process to collect Falcon 9 historical launch records from a Wikipedia page titled `List of Falcon 9 and Falcon Heavy launches`

[hide] Flight No.	Date and time (UTC)	Version, Booster [b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
1	4 June 2010, 18:45	F9 v1.0[7] B0003.1[8]	CCAFS, SLC-40	Dragon Spacecraft Qualification Unit		LEO	SpaceX	Success	Failure[9][10] (parachute)
	First flight of Falcon 9 v1.0.[11] Used a boilerplate version of Dragon capsule which was not designed to separate from the second stage.(more details below) Attempted to recover the first stage by parachuting it into the ocean, but it burned up on reentry, before the parachutes even deployed.[12]								
2	8 December 2010, 15:43[13]	F9 v1.0[7] B0004.1[8]	CCAFS, SLC-40	Dragon demo flight C1 (Dragon C101)		LEO (ISS)	NASA (COTS) NRO	Success[9]	Failure[9][14] (parachute)
	Maiden flight of Dragon capsule, consisting of over 3 hours of testing thruster maneuvering and reentry.[15] Attempted to recover the first stage by parachuting it into the ocean, but it disintegrated upon reentry, before the parachutes were deployed.[12] (more details below) It also included two CubeSats,[16] and a wheel of Brouère cheese.								
3	22 May 2012, 07:44[17]	F9 v1.0[7] B0005.1[8]	CCAFS, SLC-40	Dragon demo flight C2+[18] (Dragon C102)	525 kg (1,157 lb)[19]	LEO (ISS)	NASA (COTS)	Success[20]	No attempt
	Dragon spacecraft demonstrated a series of tests before it was allowed to approach the International Space Station. Two days later, it became the first commercial spacecraft to board the ISS.[17] (more details below)								



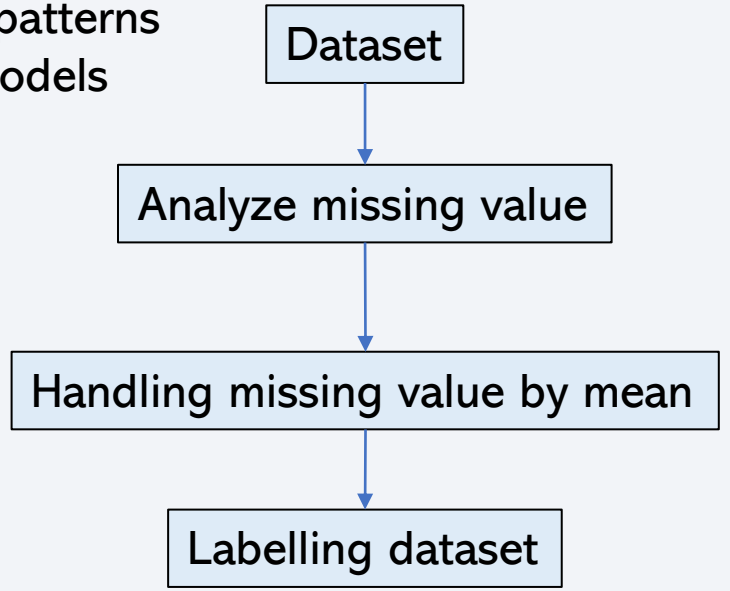
Source: Github

Data Wrangling

This process to perform some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models

```
data_falcon9.isnull().sum()
```

```
FlightNumber      0
Date              0
BoosterVersion    0
PayloadMass       5
Orbit             0
LaunchSite        0
Outcome           0
Flights           0
GridFins          0
Reused            0
Legs              0
LandingPad       26
Block             0
ReusedCount       0
Serial            0
Longitude         0
Latitude          0
dtype: int64
```

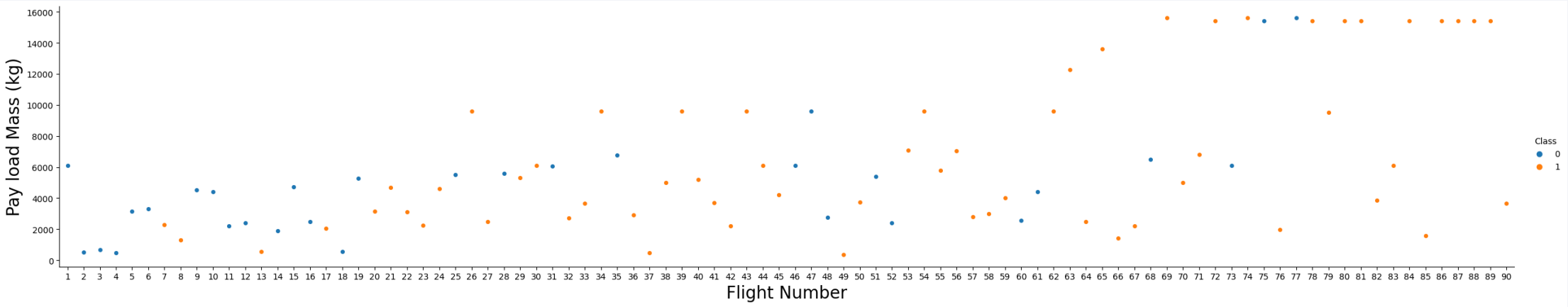


Source: Github

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
67	68	2019-08-06	Falcon 9	6500.000000	GTO	CCAFS SLC 40	None None	3	False	True	False	NaN	5.0	2	B1047	-80.577366	28.561857	0
55	56	2018-09-10	Falcon 9	7060.000000	GTO	CCAFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	5	B1049	-80.577366	28.561857	1
39	40	2017-10-11	Falcon 9	5200.000000	GTO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	3.0	1	B1031	-80.603956	28.608058	1
76	77	2020-03-18	Falcon 9	15600.000000	VLEO	KSC LC 39A	False ASDS	5	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	4	B1048	-80.603956	28.608058	0

EDA with Data Visualization

- This Process to know characteristics of SpaceX to find hidden pattern on dataset
- Chart used:
  - Scatter plot → to know relationship between 2 variables
  - Bar plot → to compare some categorical data
  - Line plot → to know trend over time



Source: Github

EDA with SQL

- This Process to answer specific question related with dataset
  - Names of unique launch sites
  - Top 5 Launch sites beginning with 'CCA'
  - Total payload mass carried by NASA (CRS)
  - Average PAYLOAD_MASS_KG_ for F9 v1.1 boosters
  - First successful ground pad landing
  - Types of boosters with successful drone ship landing and payload mass between 4.000 and 6.000 kg
  - Distinct mission outcomes and their total counts
  - Distinct booster versions with the maximum payload mass
  - Failed drone ship landings, which booster version and site, and the month of the launch
  - Rank, by count, of landing_outcomes between 2010/6/4 and 2017/3/20 in descending order

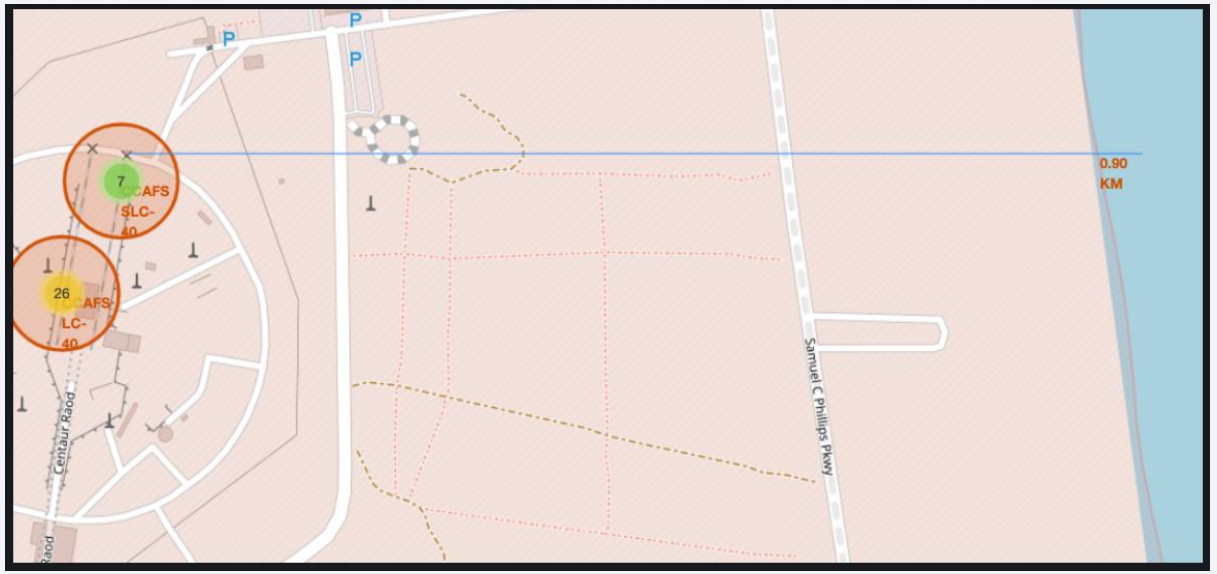
Source: Github

Build an Interactive Map with Folium

The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Finding an optimal location for building a launch site certainly involves many factors and hopefully we could discover some of the factors by analyzing the existing launch site locations.

Markers, Circles, Lines, and Marker Clusters were used in Folium Maps, visualizing launch sites and successes/failures by location as well as distances between the launch site and proximities

- Markers indicated points such as launch sites
- Circles indicated highlighted areas around specifics coordinates
- Clusters were used to indicate similar groups of events
- Lines were used to indicate and illustrate distances between two points



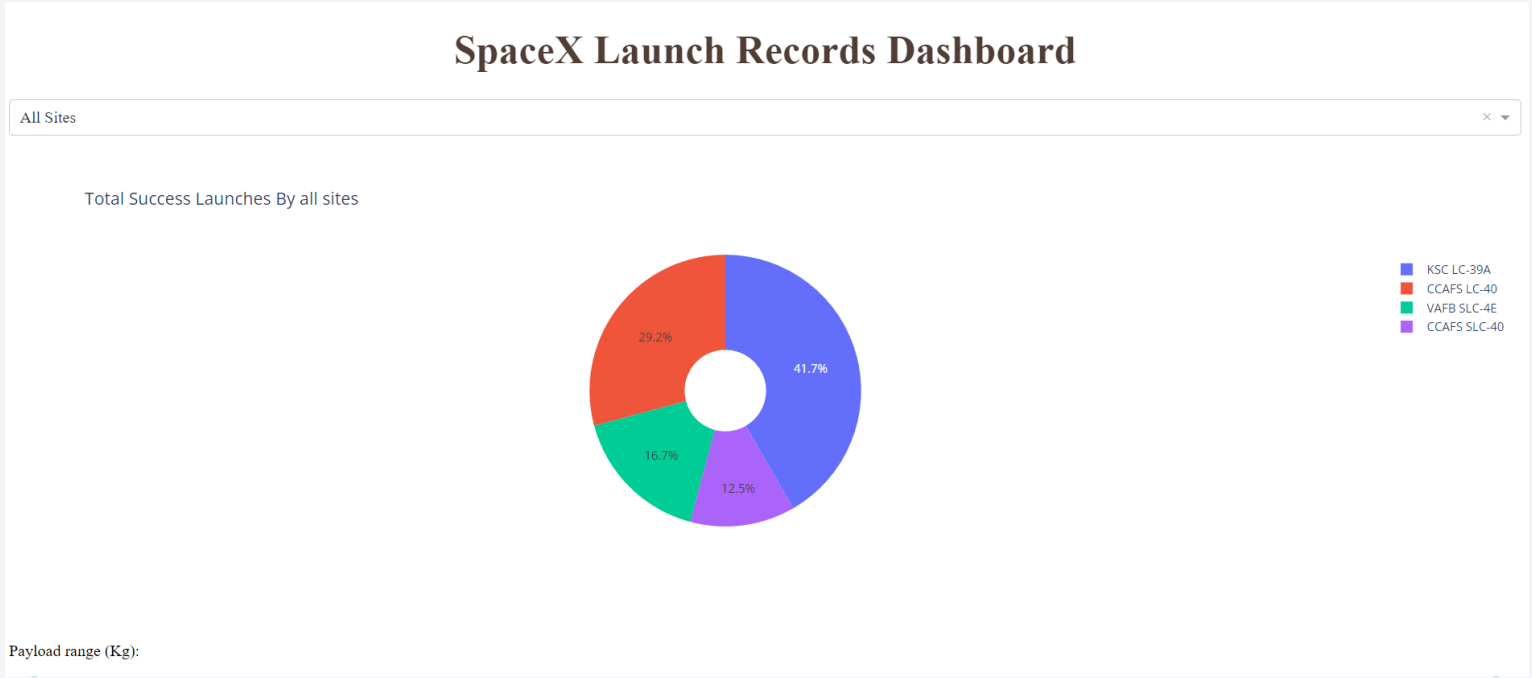
Source: Github

Build a Dashboard with Plotly Dash

To make better visualizations and make more easy to understand data, Dashboard is very useful

Chart used on Dashboard:

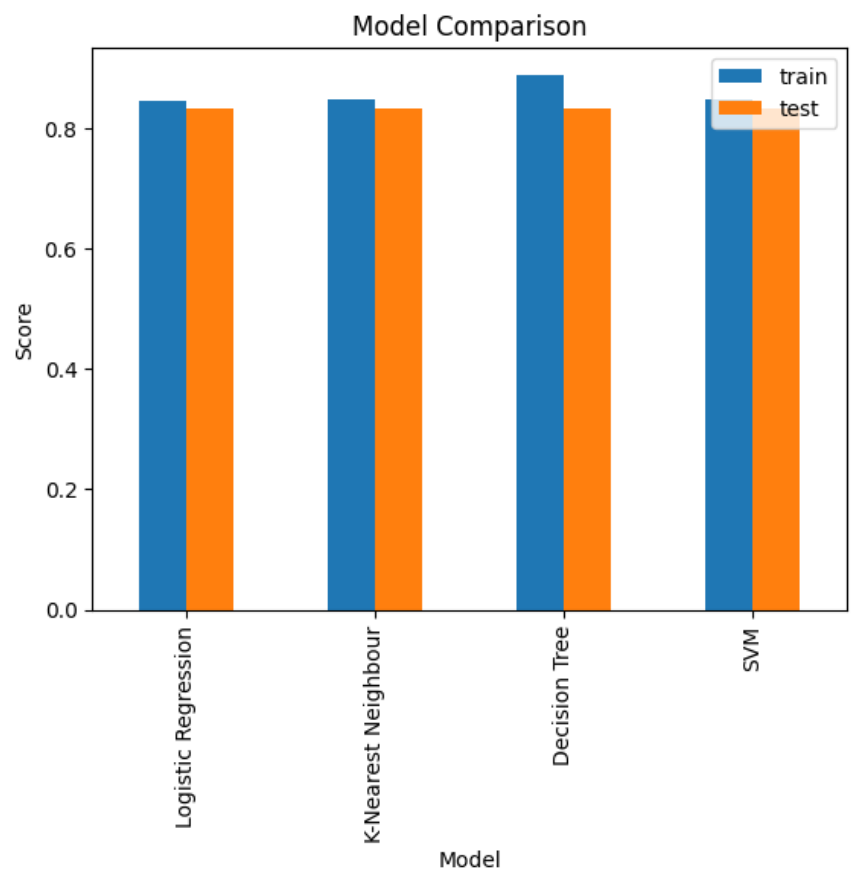
- Pie chart → compare success rate of launch site
- Scatter plot → visualize booster version by mass each class



Source: Github

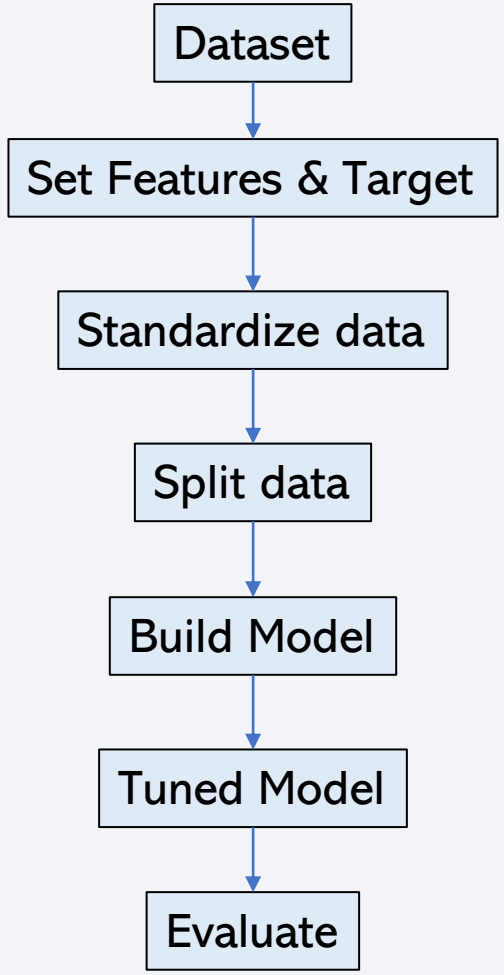
Predictive Analysis (Classification)

This Process to predict if the first stage will land given the data from the preceding labs



Name	train	test
Logistic Regression	0.846429	0.833333
K-Nearest Neighbour	0.848214	0.833333
Decision Tree	0.889286	0.833333
SVM	0.848214	0.833333

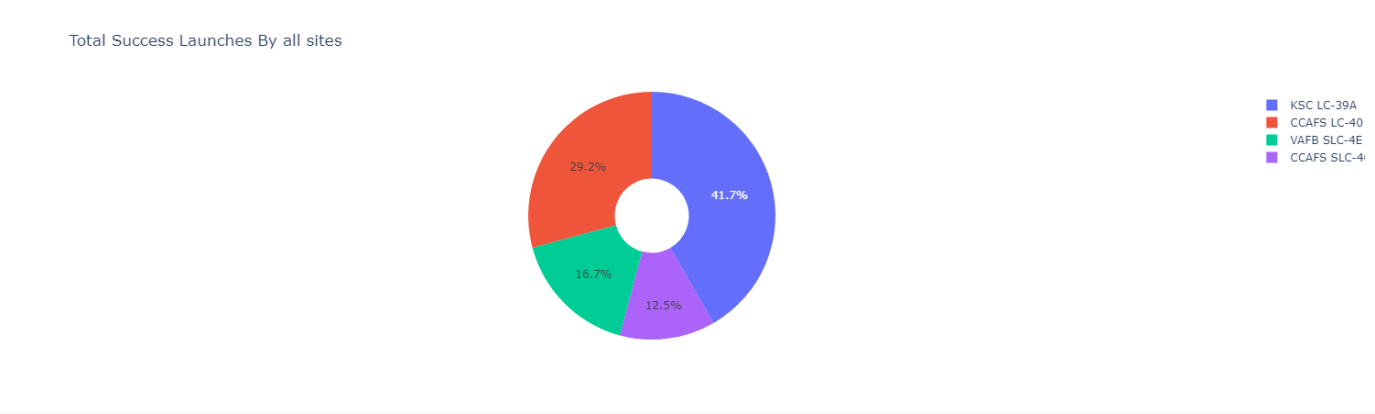
Due to all test score is same, we choose Logistic Regression because score in train and test least than others

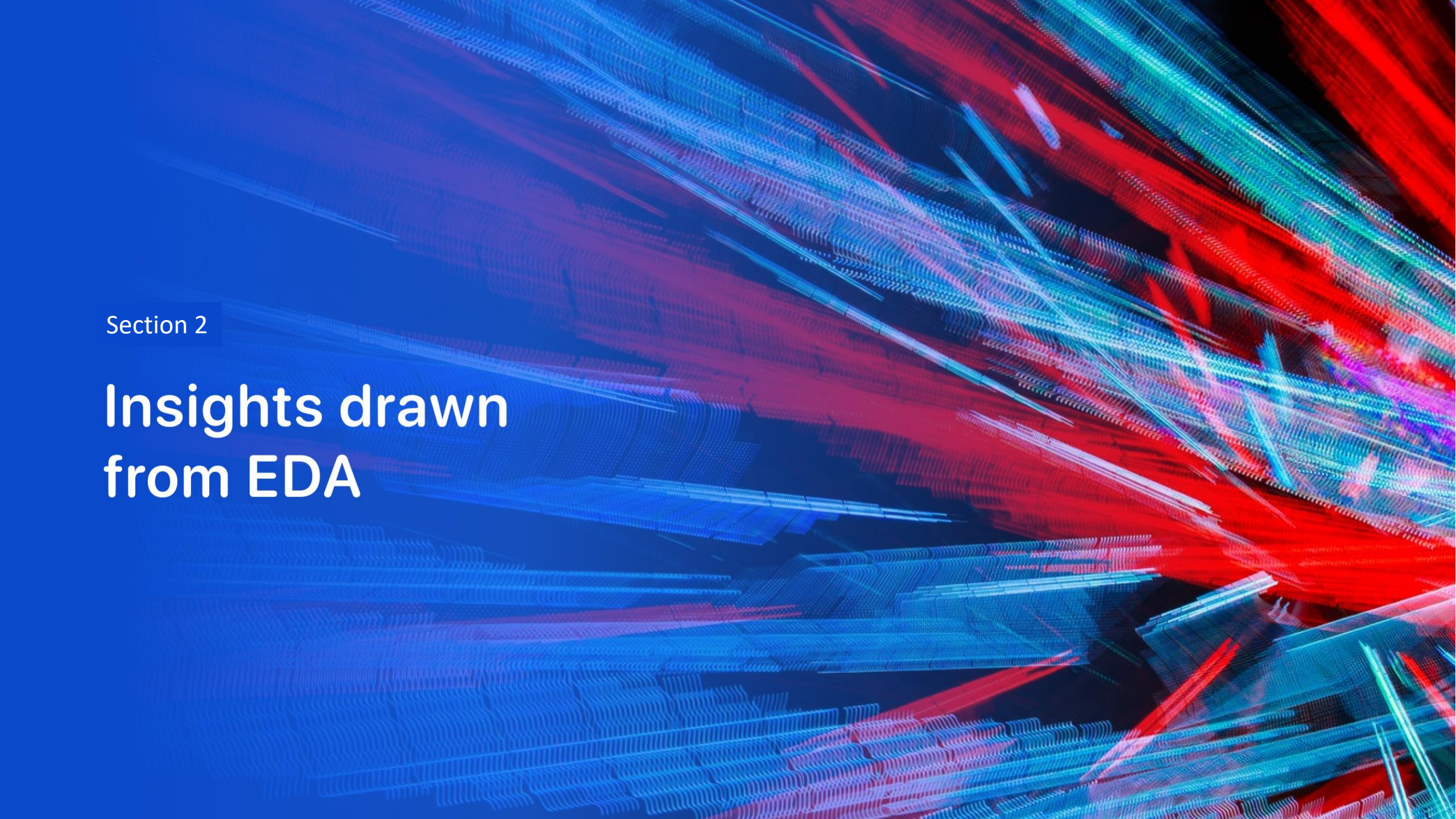


Source: Github

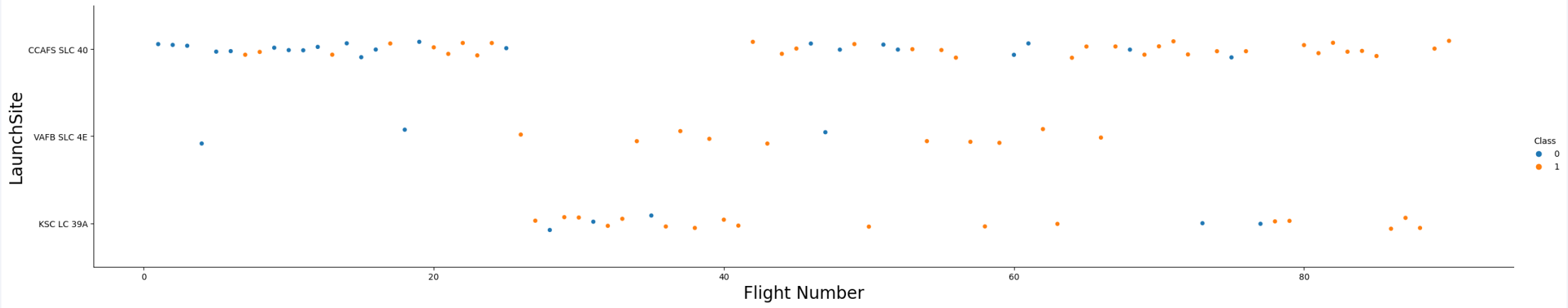
Results

- Succes rate since 2013 kept increasing till 2020
- Orbit SSO, GEO, HEO, ES-L1 with Highest success rate (100%)
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS
- Launch site CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC-4E has a success rate of 77%.
- Logistic Regression can perform better with least different score train and test, success prediction 83.3%



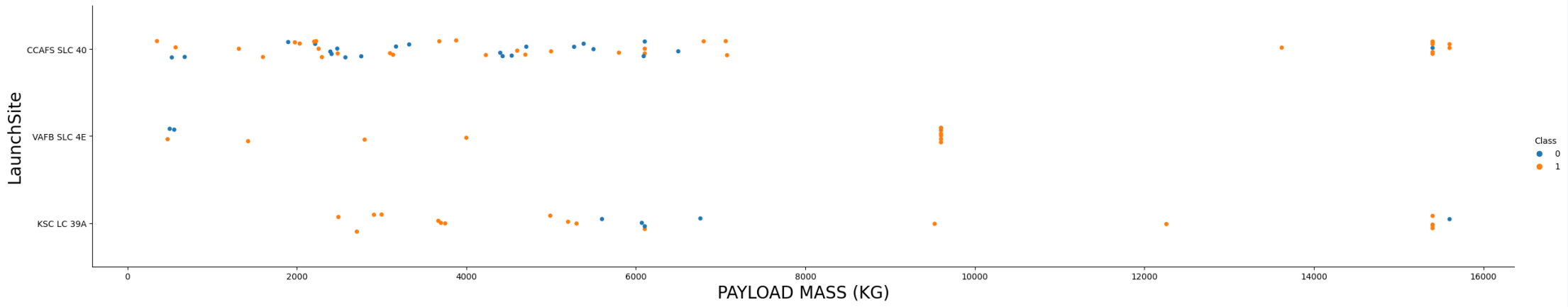

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



From 3 Launch Site in the graph, we can see in all of the launch site that mostly land successfully while the flight number increase

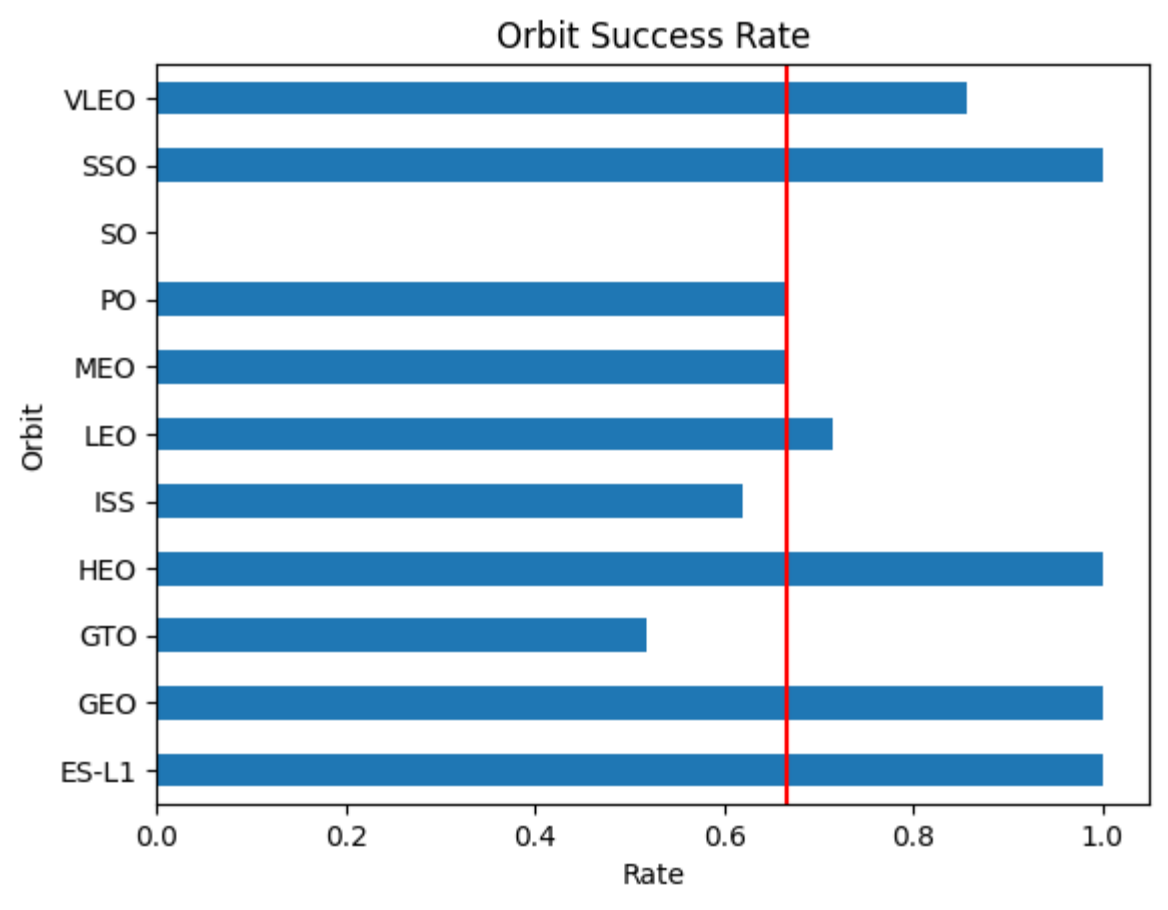
Payload vs. Launch Site



The VAFB-SLC launchsite there are no rockets launched for heavy payload mass(greater than 10.000 kg).

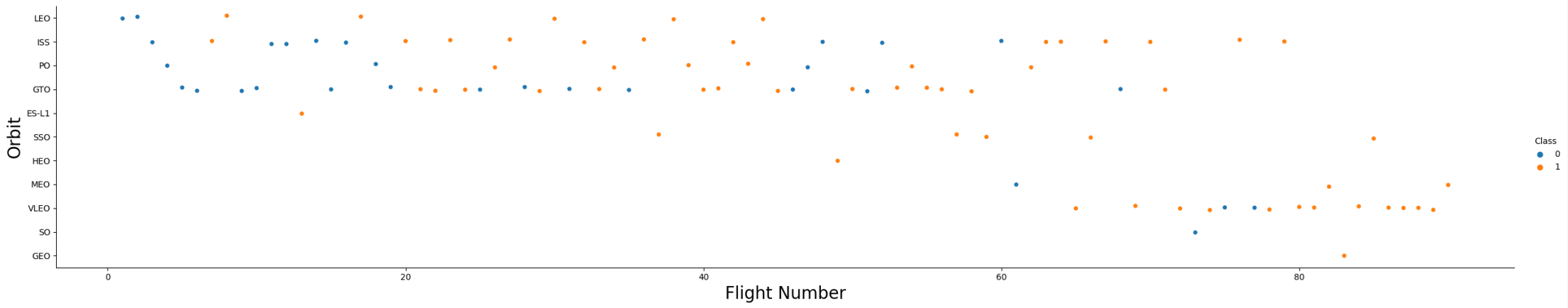
CCAFS SLC-40 Showing that mostly they didn't land successfully when payload mass below 7.000 kg

Success Rate vs. Orbit Type



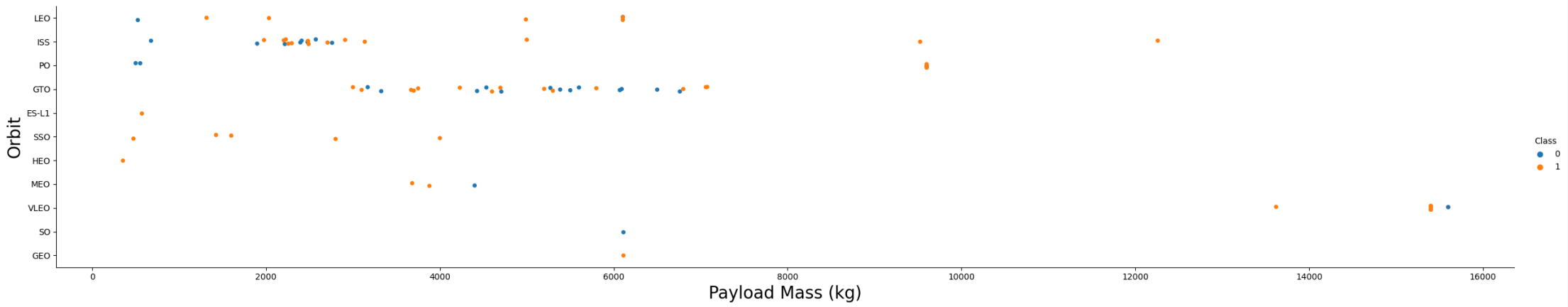
- 6/11 Orbit has success rate above average
- SO didn't record success due to only once conduct launch
- VLEO, the 3rd most orbit used, record very good success rate 85.71%

Flight Number vs. Orbit Type



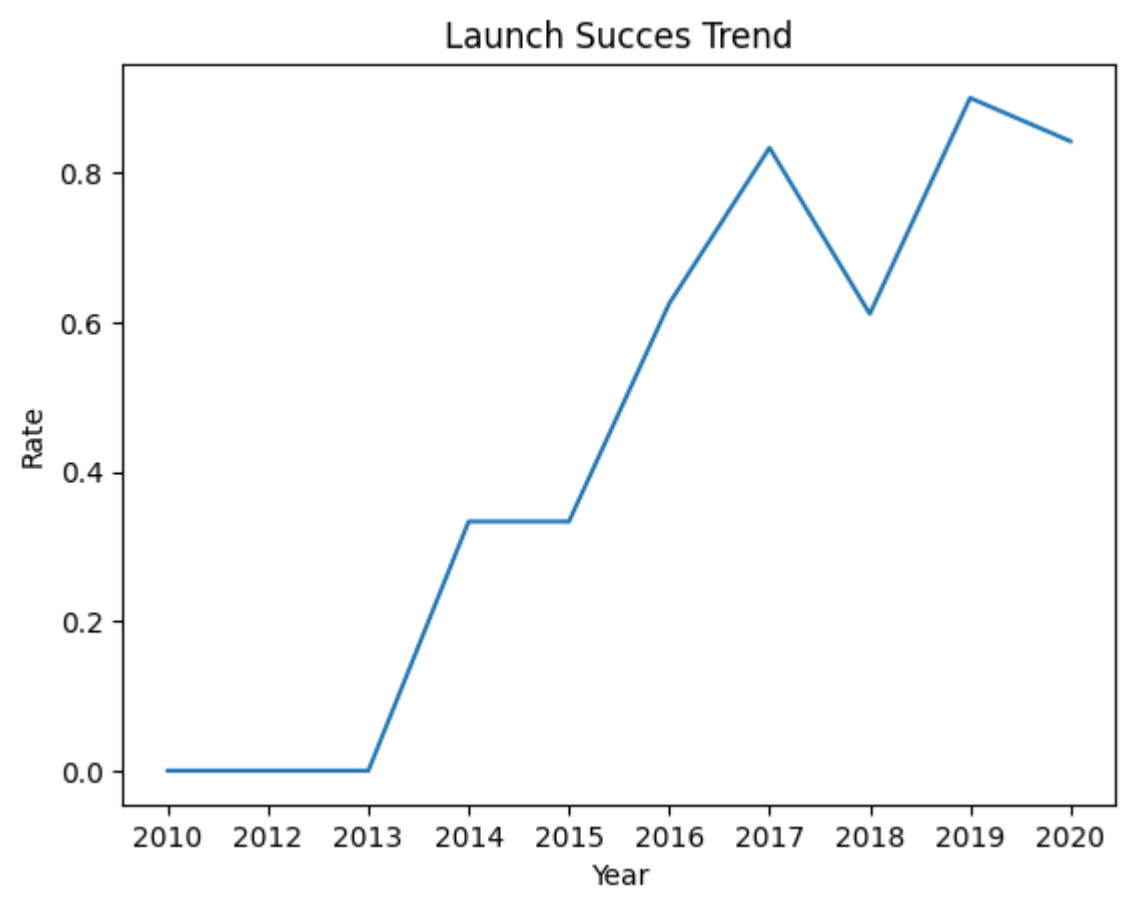
- the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



- The success rate since 2013 kept increasing till 2020
- The Highest success rate in 2019 with 90%

All Launch Site Names

```
%sql select distinct("Launch_Site") from spacextbl
```

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

- SpaceX believes that they can optimize their launch operations, and reduce launch costs, by dividing their launch missions amongst these four launch facilities: LC-39A for NASA launches, SLC-40 for United States Space Force national security launches, SLC-4E for polar launches, and South Texas Launch Site for commercial launches

Launch Site Names Begin with 'CCA'

```
%sql select * from spacextbl \
    where "Launch_Site" like 'CCA%' limit 5
```

- There are 2 launch site start with ‘CCA’, CCAFS LC-40 & CCAFS SLC-40 but only showing CCAFS LC-40 due to alphabetic order

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

```
%sql select sum("PAYLOAD_MASS__KG_") "Total_Payload_Mass" from spacextbl \
    where "customer" = 'NASA (CRS)'
```

Total_Payload_Mass
45596

- total payload carried by boosters from NASA (CRS) is the 3rd Highest with 45.596 below SpaceX and Iridium Communications

Average Payload Mass by F9 v1.1

```
%sql select Avg("PAYLOAD_MASS__KG_") from spacextbl \
    where "Booster_Version" like 'F9 v1.1'
```

Avg(PAYLOAD_MASS__KG_)
2928.4

- The average payload mass carried by booster version F9 v1.1 is 2.2928,4 Kg
- The highest average payload mass carried by booster version F9 B5 B1060.3 and the lowest F9 v1.0 B0004

First Successful Ground Landing Date

```
%sql select min("Date") from spacextbl \
    where "Landing_Outcome" = 'Success (ground pad)'
```

min(Date)
2015-12-22

- During the 2015 launch hiatus, SpaceX requested regulatory approval from the FAA to attempt returning their next flight to Cape Canaveral instead of targeting a floating platform in the ocean. The goal was to land the booster vertically at the leased Landing Zone 1 facility—the former Launch Complex 13 where SpaceX had recently built a large rocket landing pad.[108] The FAA approved the safety plan for the ground landing on December 18, 2015.[109] The first stage landed successfully on target at 20:38 local time on December 21 (01:38 UTC on December 22)

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select "Booster_Version" from spacextbl \
    where ("payload_mass__kg_" between 4000 and 6000) and ("Landing_Outcome" = 'Success (drone ship)')
```

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- Falcon 9 Full Thrust (or sometimes called Falcon 9 version 1.2) was the first version of the Falcon 9 to successfully land. Changes included a larger fuel tank, uprated engines and supercooled propellant and oxidizer to increase performance

Total Number of Successful and Failure Mission Outcomes

```
%sql select "Mission_Outcome, count("Mission_Outcome") qty from spacextbl \
    group by "Mission_Outcome"
```

Mission_Outcome	qty
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- Rockets from the Falcon 9 family have been launched 253 times over 13 years, resulting in 251 full mission successes (99.2%), one partial success (SpaceX CRS-1 delivered its cargo to the International Space Station (ISS), but a secondary payload was stranded in a lower-than-planned orbit), and one full failure (the SpaceX CRS-7 spacecraft was lost in flight in an explosion). Additionally, one rocket and its payload AMOS-6 were destroyed before launch in preparation for an on-pad static fire test. The active version, Falcon 9 Block 5, has flown 190 missions, all full successes

Boosters Carried Maximum Payload

```
%sql select "Booster_Version" from spacextbl \
    where "PAYLOAD_MASS__KG_" = (select max("PAYLOAD_MASS__KG_") from spacextbl)
```

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

- The maximum payload mass itself is 15.600 kg
- Block 5 is the latest iteration of the Falcon 9 and Falcon Heavy boosters. Changes include a stronger heat shield, upgraded engines, new carbon composite sections (landing legs, engine sections, raceways, RCS thrusters and interstage), retractable landing legs, titanium grid fins, and other additions that simplify refurbishment and allow for easier reusability.

2015 Launch Records

```
%sql SELECT substr("Date", 4, 1) "Month","Landing_OUTCOME","BOOSTER_VERSION","LAUNCH_SITE" from SPACEXTBL \
    where substr("Date", 1, 4)='2015' and "Landing_Outcome" = 'Failure (drone ship)'
```

Month	Landing_Outcome	Booster_Version	Launch_Site
5	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
5	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- These boosters looked very different from the more recent models. They were much smaller and had much less power. None of these boosters were recovered or survived landing after an orbital launch

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

```
%sql select "Landing_Outcome", count("Landing_Outcome") "Total" from spacextbl\
    where '2010-06-04'<"Date"<'2017-03-20'\
    group by "Landing_Outcome"\
    order by "Total" desc
```

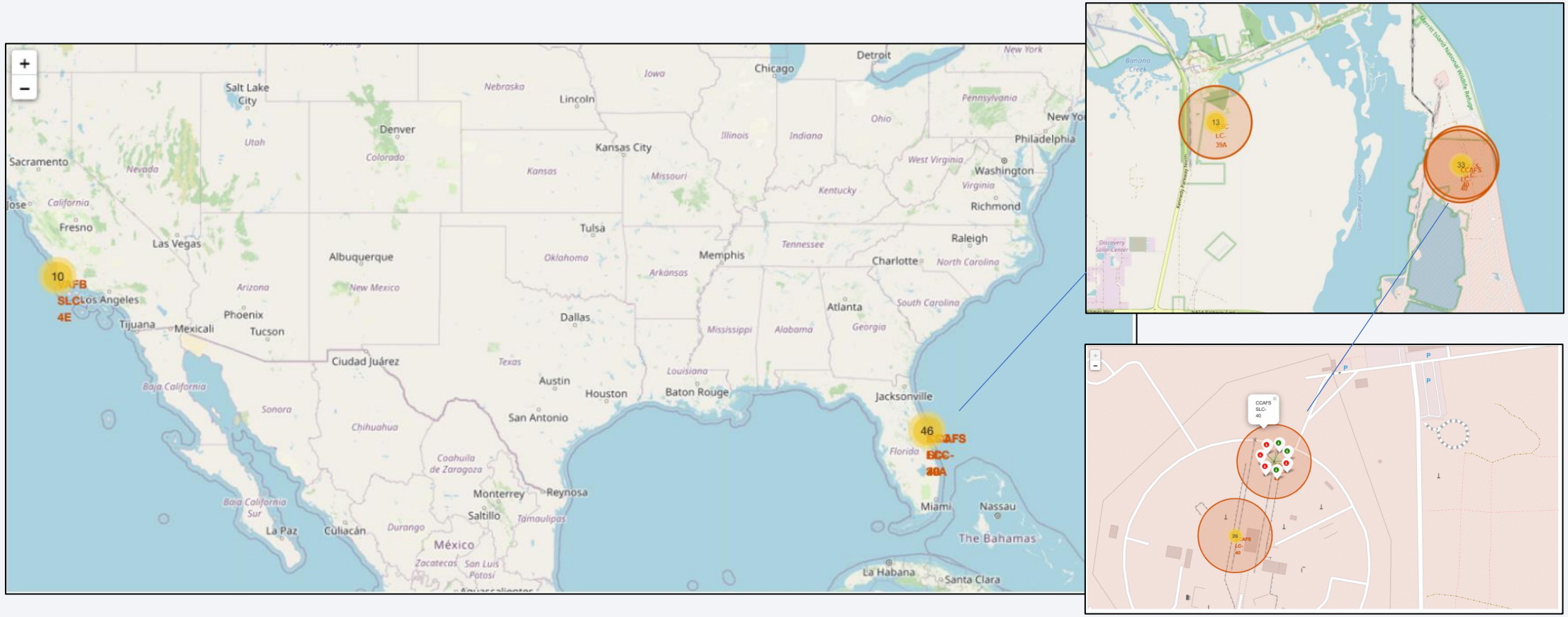
Landing_Outcome	Total
Success	38
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Failure	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1
No attempt	1

- From these date we know that falcon 9 in v1.0, v1.1 and Block 4 where the v1.0 and v1.1 none of these successfully land

Section 3

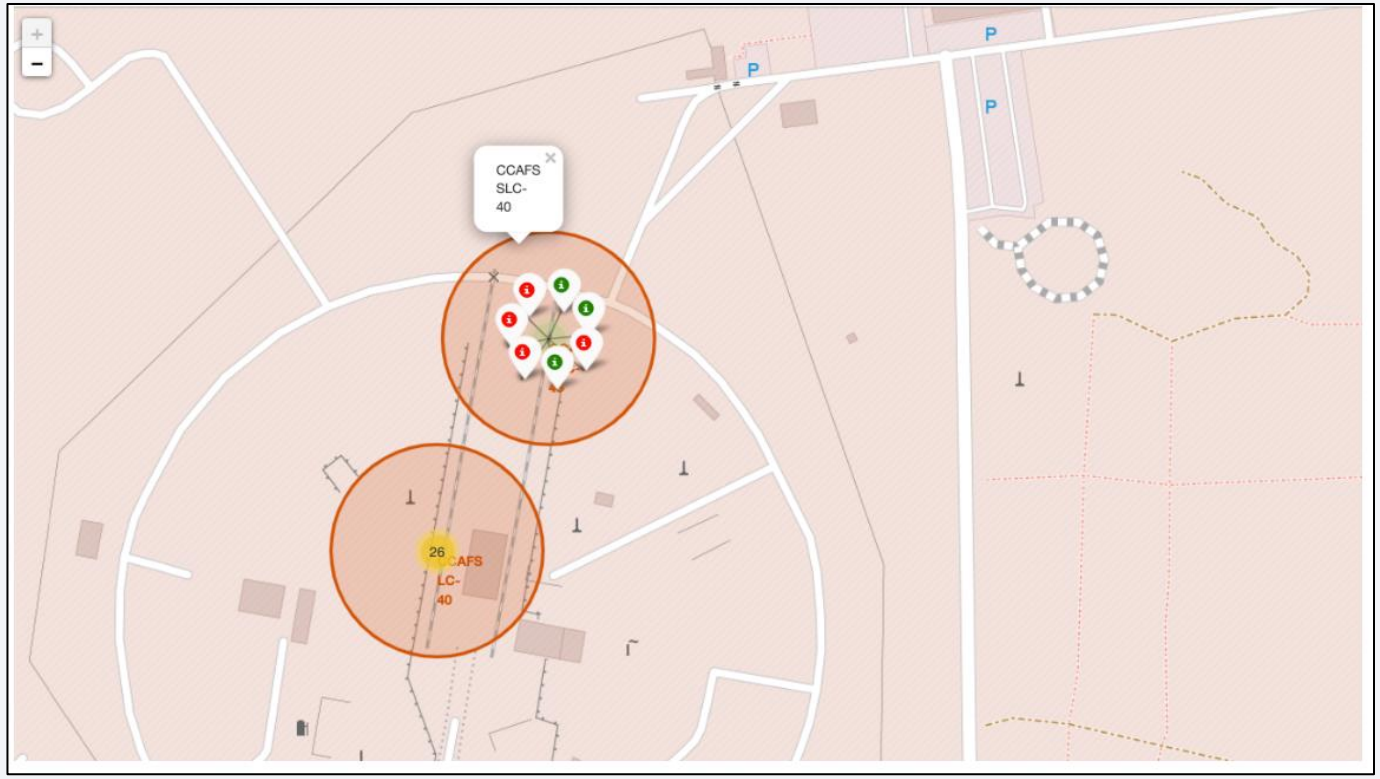
Launch Sites Proximities Analysis

Falcon 9 Launch Site Location



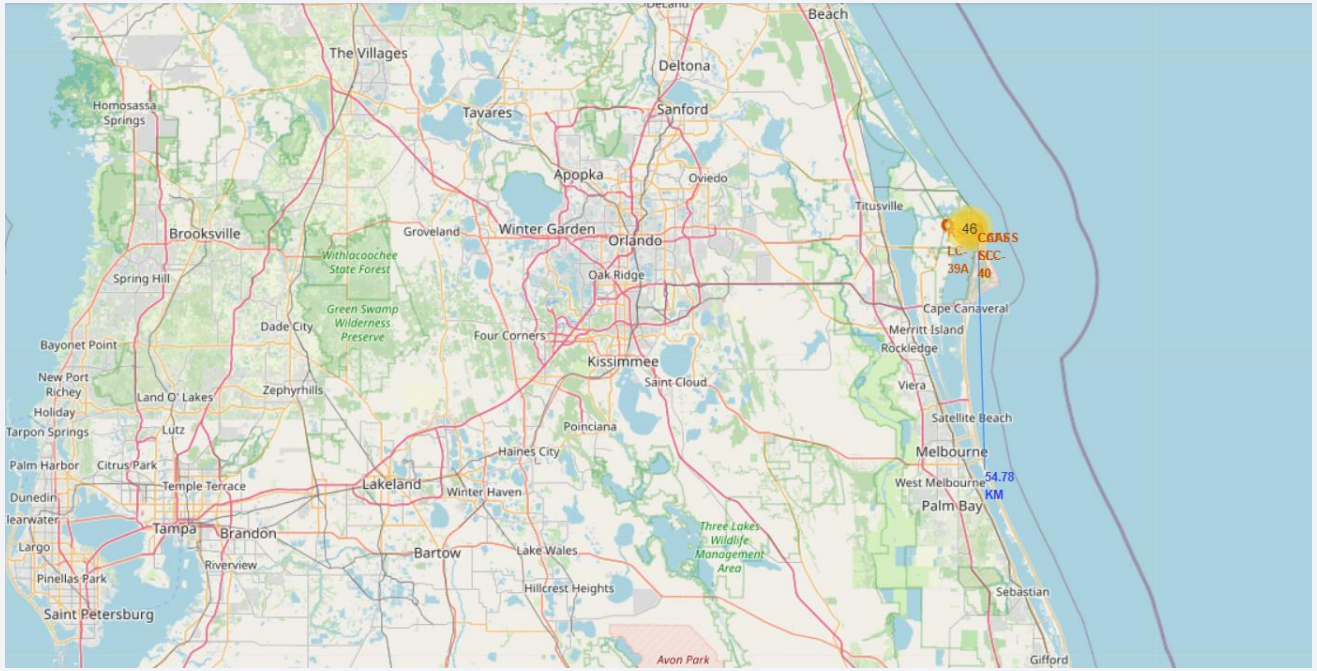
SpaceX believes that they can optimize their launch operations, and reduce launch costs, by dividing their launch missions amongst these four launch facilities: LC-39A for NASA launches, SLC-40 for United States Space Force national security launches, SLC-4E for polar launches, and South Texas Launch Site for commercial launches

Falcon 9 Launch Outcome by Site

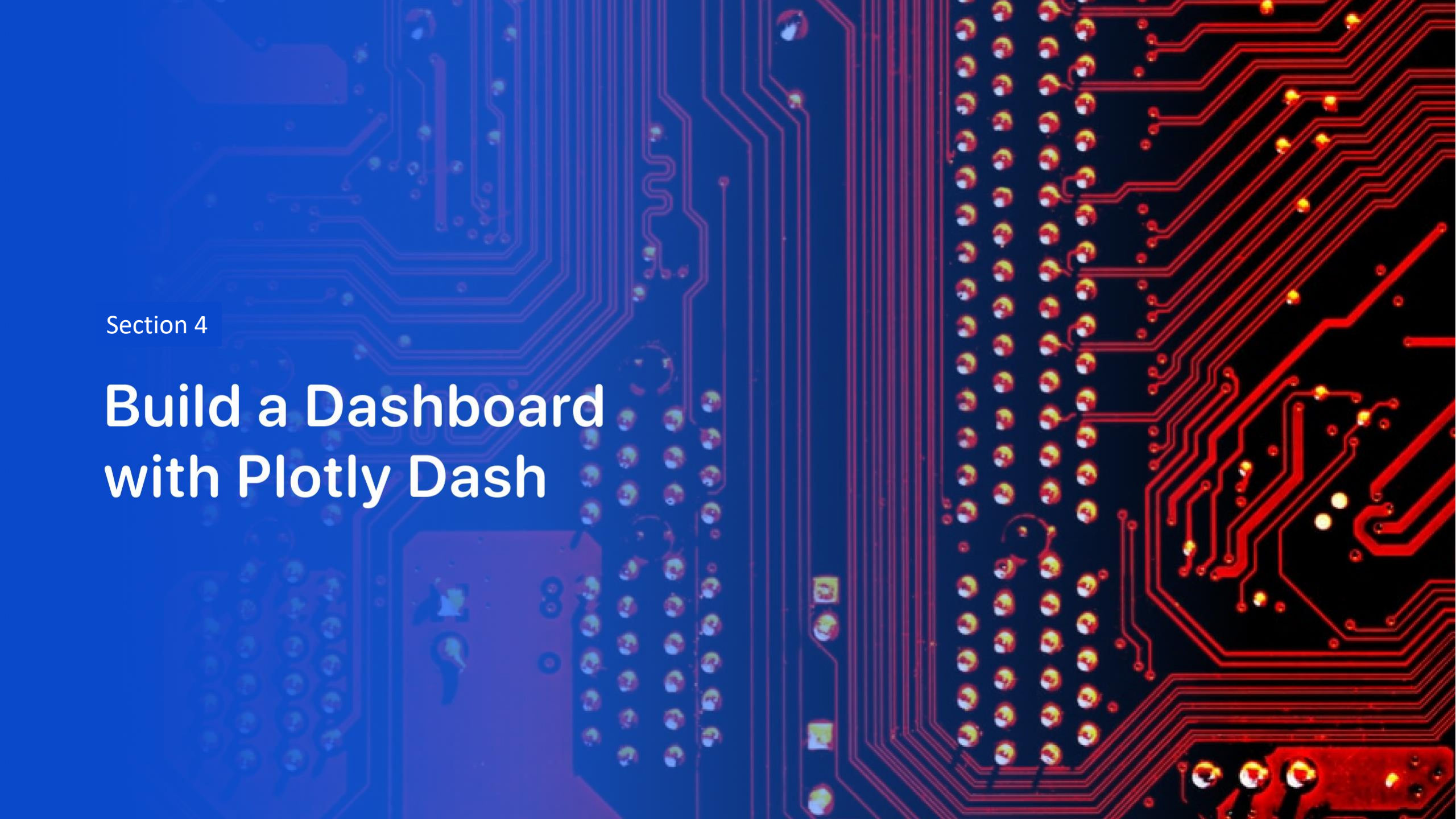


We can see the total counts by area, further refined by specific site, and then further displayed into successes (green) and failures (red).

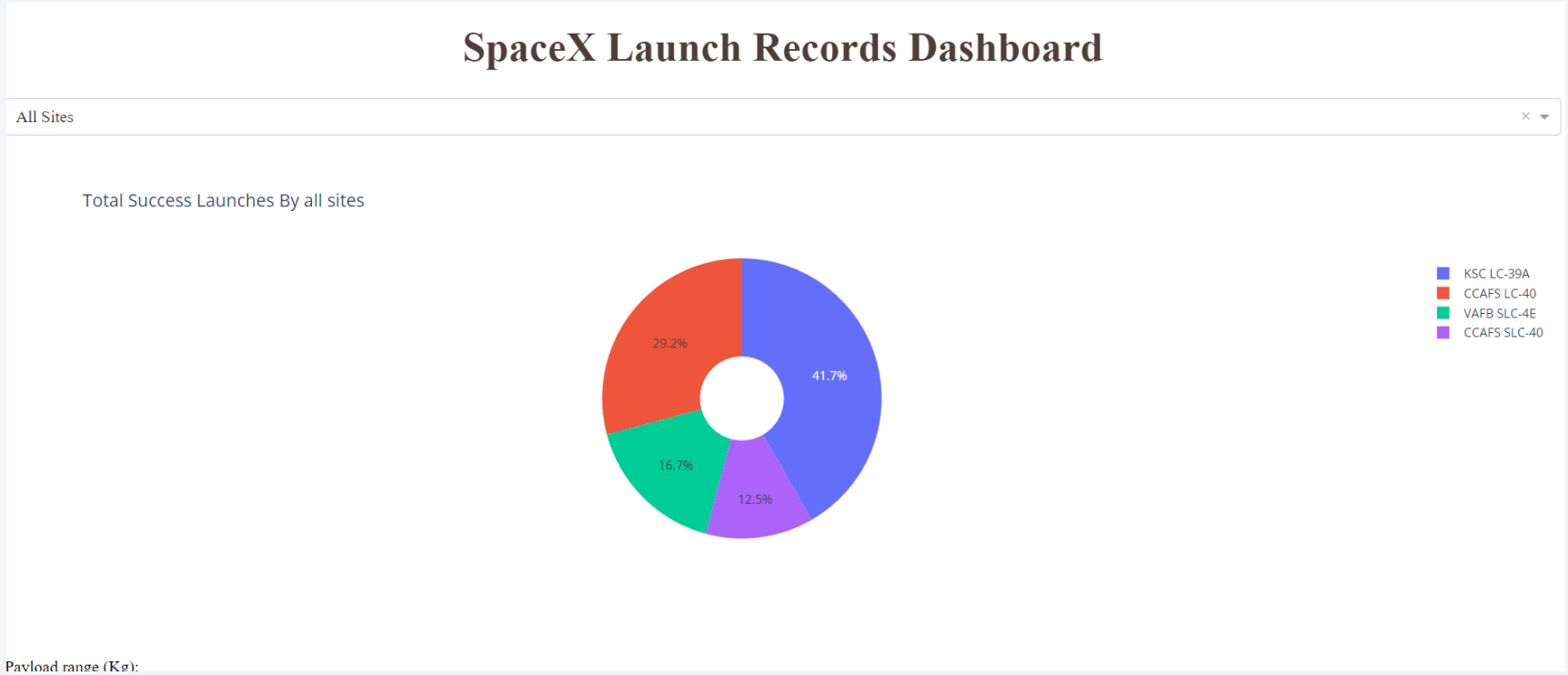
Proximity and Safety



Using various markers and searching for landmarks and their coordinates, we can do some simple math and drawing to determine the proximity of various key points. We can see that the site is close to forms of transportation, such as highways, railways, and waterways, while being much farther away from more inhabited areas.

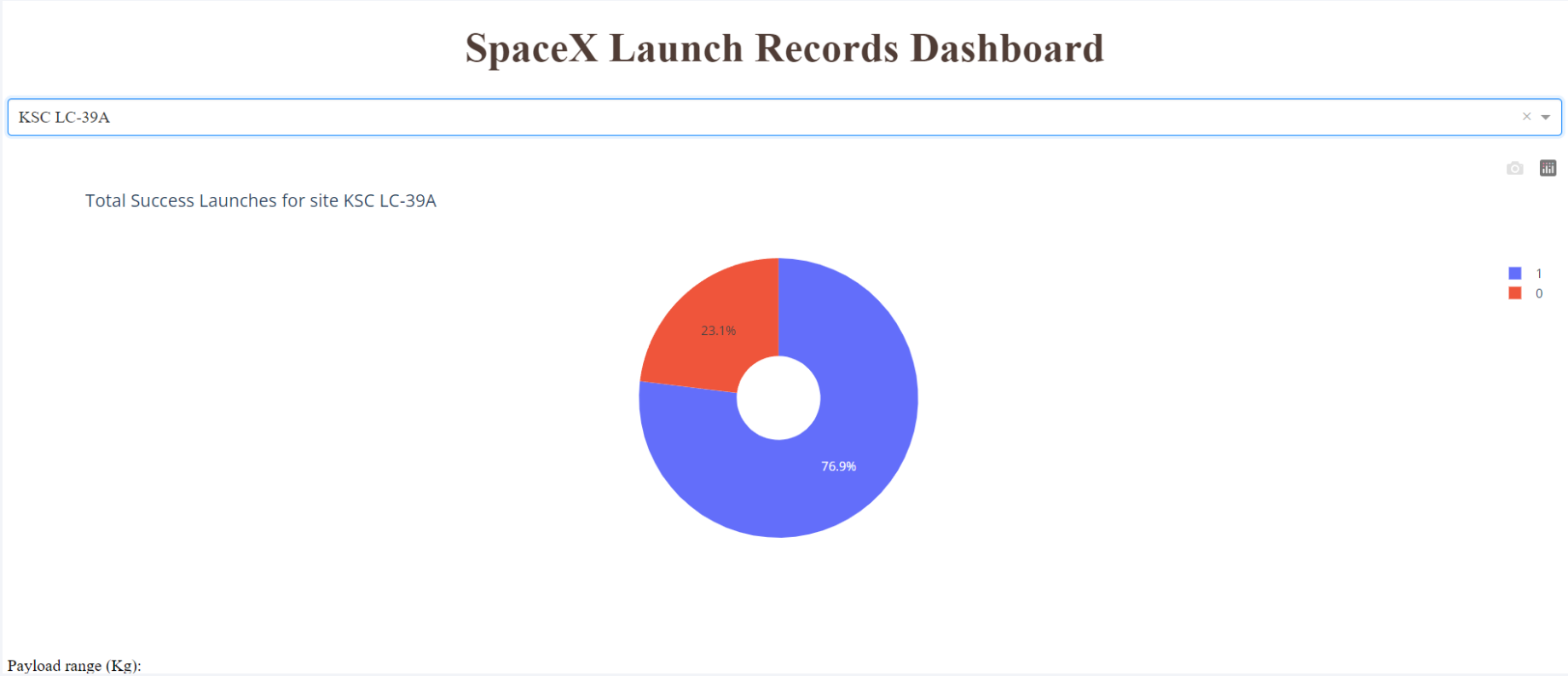

Section 4
Build a Dashboard with Plotly Dash

SpaceX Success Rate by Launch Site



- The Highest Succes rate Launch Site is KSC LC-39A
- In 2015, SpaceX built the Horizontal Integration Facility (HIF) just outside the perimeter of the existing launch pad in order to house both the Falcon 9 and the Falcon Heavy rockets, and their associated hardware and payloads, during preparation for flight.[20] Both types of launch vehicles are transported from the HIF to the launch pad aboard a Transporter Erector (TE) which rides on rails up the former crawlerway path

KSC LC-39A Launch Site



- Success Rate this Launch site is 77%
- The first SpaceX launch from pad 39A was SpaceX CRS-10 on February 19, 2017, using a Falcon 9 launch vehicle; it was the company's 10th cargo resupply mission to the International Space Station,[32] and the first uncrewed launch from 39A since Skylab

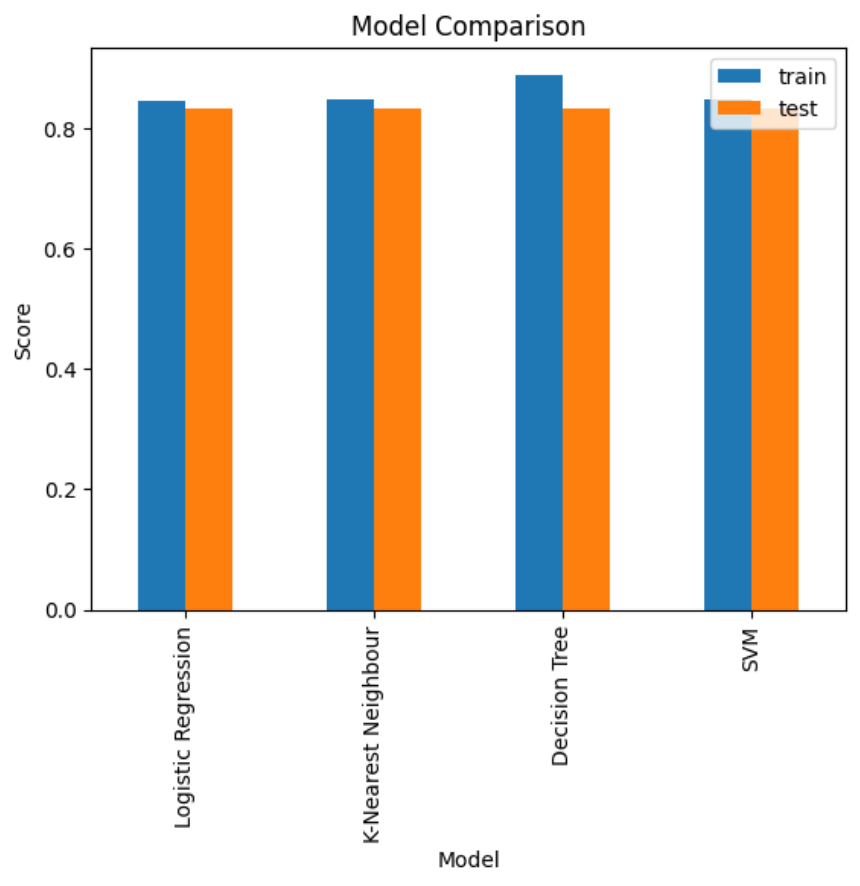
Booster Version Payload Mass

- From Range 2.000 to 8.000 Kg the most weighest Booster is F9 FT B1037 which is Block 4 wigh 6.761 Kg but didn't land successfully

Section 5

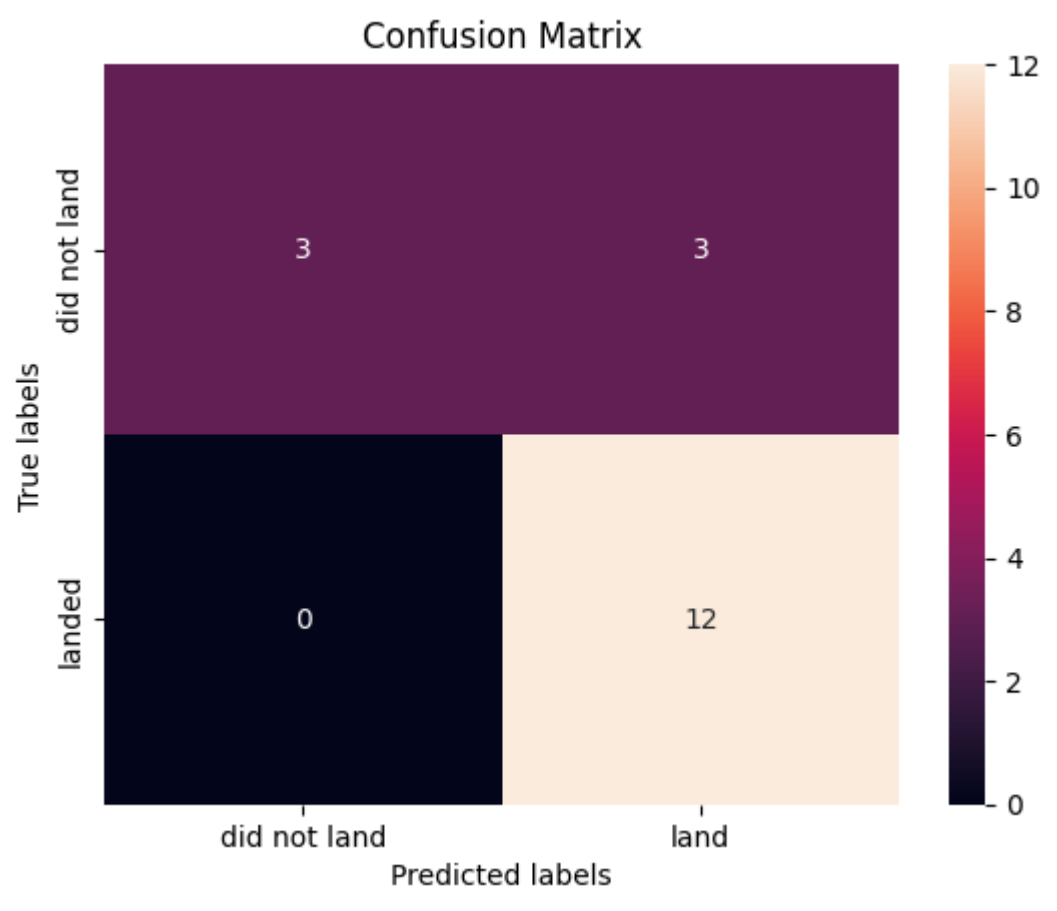
Predictive Analysis (Classification)

Classification Accuracy



- The Highest accuracy using Decision Tree Model with 88.92%

Confusion Matrix



- From these model we know that model can predict 100% by TPR
- And FPR of this models is 50%

Conclusions

- SpaceY can reduce cost by predicting first stage of launching with accuracy 83.3% using Decision Tree Classifier
- There are 4 Launch site used by SpaceX with the highest success rate is KSC LC-39A
- SpaceY can identify optimal launch site location using Folium Interactive Map
- Success Rate of Rocket Launch increasing year by year, SpaceY can adapt concept from SpaceX

Appendix

- All Source can be viewed here:
  - https://github.com/aliyafi94/Coursera_IBM-Data-Science-Assignment/tree/main/10.%20Capstone%20Project

Thank you!